

SEGRADA: EINE SEMANTISCHE GRAPHDATENBANK ALS WERKZEUG FÜR CITIZEN SCIENCE

Maximilian Kalus

Keywords

Datenbank, Semantik, Graphdatenbank, Geschichte, Quellenreferenzierung

Zusammenfassung

Segrada, eine Graphdatenbank zur Aufbereiten geisteswissenschaftlicher, semantischer Informationen, versteht sich als Werkzeug der wissenschaftlichen Arbeit: Gerade in der Geschichte und anderen Geisteswissenschaften ist das Erfassen von Daten aus verschiedenen Quellen oft zeitaufwändig, fehleranfällig und für Dritte nur schwer nachzuvollziehen. Segrada bricht die Informationen in kleinere Einheiten auf, welche untereinander verknüpft werden können. Die Daten liegen daher in einer formal eindeutigeren Weise vor, welche sowohl von Maschinen als auch von Menschen leicht lesbar ist. Die Validität der Datensätze wird durch Quellenangaben gewährleistet. Außerdem können die Quellen selbst in der Datenbank gespeichert und durchsuchbar gemacht werden. Da gerade in den Geisteswissenschaften Daten oft bereits in ihren Quellen ungenau sind, erlaubt die Datenbank alternative Schreibweisen von Namen und Orten, aber auch ungenaue Datumsangaben. Darüber hinaus kann Segrada dank der Validierung der Quellen und Daten sowie dank des Hinterlegens mit Informationen zu demjenigen, der die Daten einspeist, als Hilfswerkzeug zur systematischen Erfassung und Aufbereitung von Daten aus Citizen Science-Projekten dienen, deren Ursprung jederzeit zurückverfolgt werden kann.

Auf Basis einer älteren Forschungsdatenbank entwickelte der Autor Segrada. Personen, Orte, Ereignisse und andere Informationseinheiten werden als „Knoten“ in der Datenbank erfasst und untereinander durch semantische Verbindungen verknüpfbar. Sowohl Knoten als auch Verknüpfungen, Kanten genannt, dürfen dabei mit Geo- und Zeitdaten (Perioden bzw. Zeitpunkte) versehen werden. Als technische Basis dient eine einfach zu startende Webapplikation auf Java-Basis. Sie kann als Programm, als Server, in einem Serverumfeld oder als Docker-Container ausgeführt werden. Dadurch ist der Einsatz sowohl für Einzelpersonen als auch für Gruppen möglich. Die Anwendung ist plattformunabhängig und läuft unter Windows, OSX, Linux und BSD. Das Programm ist unter der Apache Open Source-Lizenz veröffentlicht und der Source Code auf Github einsehbar.

Abstract:

First and foremost, Segrada is a scientific tool: In historical research and other humanities, data aggregation is tedious, error prone and hard to reproduce later on. The database enables the researcher to gather data incrementally into formal bits of information. Consequently, data is defined more rigidly and computable by both machines and humans. Bits of data can be annotated by source references in order to validate these. Moreover source documents can be saved into the database and are full text searchable, like the rest of the database. Fuzzy data is a constant challenge in the humanities. Sources can be ambiguous, spellings can differ. Segrada compensates this need by allowing to add alternative spellings and fuzzy time data. Moreover, thanks to the validation of sources and data, and thanks to the rear fastening with information about the person who feeds the data, Segrada can serve as auxiliary tool for the systematic collection and processing of data from citizen science projects, whose origin can always be traced.

Based on an earlier research database, the author of this paper has developed Segrada, a graph database to process semantic information in the humanities. People, locations, events and other pieces of information can be created as “nodes” within the database that can be linked to each other. Different connection types are possible. Both nodes and connections, called or edges of the graph, can contain additional geographic and time data. The technical basis of Segrada is an easy to start Web-based Java application. The application can be run on the desktop, as a standalone server, within a Servlet or as Docker container. Consequently, Segrada can be used by single researchers or by whole groups. The application is not platform dependent and can be run on Windows, OS X, Linux or BSD. It has been published under the Apache open source license with the source code accessible on Github.

Einleitung

Historiker und Bürgerforscher teilen sich so manche Gemeinsamkeit: Die Recherche ist arbeitsreich und dreht sich um die kleinen, aber wichtigen Details. Dabei gewonnene Informationen sind oftmals ungenau, etwa bei der Schreibweise von Personennamen, und widersprüchlich, gleichzeitig jedoch komplex. Auch können die Zeitangaben der Zeugen schwammig sein – wie glaubwürdig ist „so gegen 10 Uhr“? Eine weitere Komponente ist die zeitlich-örtliche Ebene, die essentiell ist, wenn es darum geht herauszufinden, wo sich welcher Akteur wann aufgehalten hat. Die Anwen-

dung ist daher nicht nur für Forscher interessant, sondern auch für jeden Laien, der Informationen strukturieren will.

Traditionellerweise haben Historiker und auch Bürgerforscher solche Informationen – oft unabhängig voneinander – in ihren Köpfen bewahrt oder aber in Form von Notizbüchern, Karteikarten, Akten oder kopierten und glossierten Buchauszügen. Während diese Aufbewahrungsmittel und -konzepte ohne Zweifel ihre Vorzüge besitzen, haben sie auch drei gravierende Nachteile: Zunächst fällt es dem Benutzer schwer, einmal angelegte Daten schnell zu durchsuchen. Selbst ein gutes Karteikartensystem mit Referenzierung auf andere Karten ist mühselig in seiner Benutzung. Des Weiteren ist die Übertragbarkeit der Daten an andere Personen nicht ohne Weiteres möglich. Die Informationen müssen oftmals händisch kopiert werden, was das Arbeiten in einem Team erschwert. Als letzter Nachteil ist die Nachvollziehbarkeit der Daten zu nennen. Die meisten Ablagesysteme sind zunächst nur für den Ersteller selbst verständlich, selbst wenn sie standardisierten Karteisystemen oder dem Prinzip eines Zettelkastens nach Niklas Luhmann folgen. Dies ist insofern kritisch, weil auch die Geisteswissenschaften sich zunehmend Fragen nach der Offenlegung von Rohdaten stellen müssen, wie es im naturwissenschaftlichen Bereich bereits üblich ist. Wissenschaftliche Rohdaten erlauben es nicht nur, die aus der Auswertung entstandenen Thesen zu überprüfen, sondern auch, die Daten später für neue Forschungsansätze weiter zu nutzen. Die klassische Veröffentlichung in Form einer Monographie oder eines Artikels als Endprodukt wissenschaftlichen Schaffens wird damit zunehmend in Frage gestellt. Denn solche Publikationen sind zwar angenehm zu lesen und nachzuvollziehen, bieten hingegen wenig Möglichkeiten zur Überprüfung der Aussagen. In Zeiten eines immer rasanter werdenden Wissensaustausches, von Big Data und gewissen Verdrängungseffekten im Wissenschaftsbetrieb ist grundsätzlich die Neuinterpretation von Quellen nicht in Frage zu stellen. Vielmehr geht es darum, die Hermeneutik und andere genuin historische Methoden in Aspekten zu ergänzen und moderne Hilfsmittel einzusetzen, um den wissenschaftlichen Prozess zu verbessern und nachvollziehbarer zu machen.

Die Suche nach der perfekten Forschungshilfe

Bei obigen Überlegungen können elektronische Hilfsmittel die Anforderungen an Durchsuchbarkeit, Übertragbarkeit und Nachvollziehbarkeit deutlich besser bedienen als nicht-elektronische Arten der Informationserschließung. Zunächst sind elektronische Varianten nicht-elektronischer Datenablagemöglichkeiten zu nennen. Klassischerweise wären dies das Dateisystem eines modernen Computers. Daneben gibt es Software zur Abbildung von Karteien oder Wissensseinheiten, beispielsweise das Programm *Zettelkasten*, welches das System des bereits erwähnten Niklas Luhmann ins elektronische Format überträgt. Auch Programme wie *Evernote*, *Wunderlist* oder *The Brain* gehören in diese Kategorie. Bei solchen Systemen fehlt aber oftmals die Verknüpfung von Daten, die einfache Durchsuchbarkeit oder die Möglichkeit zur Zusammenarbeit mit anderen Menschen. Alternative Software zur archivalischen, objektbezogenen oder geographischen Erfassung von Daten existiert bereits seit geraumer Zeit. Zu nennen sind hier beispielsweise *FAUST* für Archive oder diverse GIS-Anwendungen zur Verarbeitung geographischer Daten. *UCINET*, *Pajek* und *Gephi* sind Vertreter Netzwerk-basierter Software. Solche Programme sind aber generisch oder für spezielle Einsatzzwecke geschaffen, also auf spezifische Fragestellungen zugeschnitten, beispielsweise die Erfassung ortsbezogener Daten oder die Analyse von Netzwerken. Für eine umfassende Nutzung und Erfassung historischer Daten eignen sie sich oftmals weniger. Außerdem fehlt der Mehrheit dieser Programme die Möglichkeit zur kollaborativen Zusammenarbeit (gemeint ist hier E-Collaboration, also die Zusammenarbeit über ein Netzwerk verschiedener Endgeräte), weil diese Anwendungen jeder Nutzer individuell und „offline“ auf seinem Rechner installiert. Ähnlich sieht es im Feld der Forschungsdatenbanken aus. Meistens handelt es sich dabei um die Früchte von Forschungsprojekten, die online für Interessierte zugänglich gemacht werden. Auch hier steht ein spezieller Einsatzzweck im Vordergrund, eine Nachnutzung oder Ergänzung der Daten ist in der Regel nicht vorgesehen. Die Verwendung der darunterliegenden Software oder der Rohdaten ist in der Regel ebenfalls nicht ohne Weiteres möglich, Downloads, Installationsanleitungen oder Informationen über technische Details werden nur in Ausnahmefällen angeboten.

Kommerzielle Systeme wie *FAUST*, *Evernote* oder *Citavi* tragen zudem den Nachteil der Abhängigkeit zum Anbieter der Software. Dieser kann jederzeit die Bedingungen für sein Produkt ändern, beispielsweise welche Betriebssysteme unterstützt werden oder unter welchen Bedingungen die eigenen Daten nutzbar sind. Die Zukunftsfähigkeit des angebotenen Dienstes ist zudem aufgrund technischen Fortschritts oder des finanziellen Zusammenbruchs des Anbieters nie gesichert. Der Anbieter solcher „geschlossener“ Software hat außerdem in der Regel wenig Anreiz, die Daten auch außerhalb seines Portfolios exportierbar zu machen, was den Austausch dieser erschwert. Auch der jahrelange Streit um die Konvertierbarkeit von Microsofts Office-Formaten in das bereits länger ISO-zertifizierte und durch EU-Gesetze für öffentliche Stellen vorgeschriebene Open Document Format (ODF) zeigt, welche Probleme im Zusammenhang mit kommerziellen Anbietern entstehen können.

Dies und die immer häufiger notwendige Zusammenarbeit mit anderen Menschen legt die Nutzung von offenen Web-basierten Systemen nahe. Am bekanntesten ist sicherlich Wiki-Software, welche sich gut zur Dokumentation und Selbstorganisation eignet, wie das Beispiel *Mediawiki* für das Online-Lexikon Wikipedia zeigt. Daneben gibt es eine Reihe von offenen Lösungen für Notizbücher, Inventare, Genealogie sowie Baukästen für strukturierte Daten oder Dokumente, wie *Structr* oder *Apache Wave*. Obwohl diese Systeme für bestimmte Anwendungsgebiete gut geeignet sind, eignen sie sich nur bedingt für die eigene Forschungsdokumentation. Wikis bieten beispielsweise zwar Rückverknüpfungen an – welche Artikel verlinken auf einen anderen? –, aber eine Sortierung nach Datumsangaben oder Geographie ist meistens nicht ohne zusätzliche Plug-ins vorgesehen.

Historische semantische Netzwerke

Im Folgenden werden Ansatzpunkte zur Lösung der oben genannten Probleme vorgestellt. Dabei geht es zunächst um einen theoretischen Ansatz, den der historischen semantischen Netzwerke, gefolgt von der Referenzanwendung namens *Segrada*, einer semantischen Graphdatenbank.

Historische semantische Netzwerke fußen auf mehreren älteren Ansätzen und kombinieren diese. Zunächst ist die Prosopographie zu nennen, die Erfassung aller relevanter Daten zu einem bestimmten Personenkreis in systematischer Weise. In der Forschung wird diese Methode seit Langem genutzt (z. B. [Kirchner 1901/02](#)). Manche der Werke beziehen explizit semantische oder netzwerkartige Konzepte mit ein, beispielsweise jene von Wolfgang Reinhard ([1979, 1996](#)). Analog dazu haben Historiker früh begonnen, die aus der Soziologie kommende soziale Netzwerkforschung auf geschichtliche Fragestellungen anzuwenden. Die historische Netzwerkforschung wird in den letzten Jahren an Popularität in vielen Forschungsprojekten (z. B. [Häberlein 1998](#); [Gramsch 2013](#); [Collar 2013](#); [Preiser-Kapeller 2016](#)) zur Auswertung bereits erfasster Daten eingesetzt. Die Stärke der Methode liegt zum einen in der Möglichkeit, große Datenmengen automatisiert auszuwerten. Dazu steht eine Vielzahl an primär mathematischen Methoden zur Verfügung, um einzelne Akteure innerhalb eines Netzwerks zu verorten – beispielsweise seine Zentralität oder seine Zugehörigkeit zu einem bestimmten Cluster innerhalb des Netzwerks. Zum anderen lassen sich Netzwerke dadurch konkret visualisieren. Beide vorgestellten Ansätze tangieren das Forschungsfeld des Wissensmanagements und Wissensaustausches. Dieser Bereich war in den letzten Jahren besonders dynamisch. Vor allem Problemstellungen aus der Forschung zu Sozialen Netzwerken und dem World Wide Web haben die Entwicklung neuer Technologien und Konzepte befeuert. So versucht das sogenannte Semantic Web, globales, auch wissenschaftliches Wissen in aufbereiteten Daten zu erfassen und grafisch mit Verbindungen zu zugehörigen Informationen abzubilden ([Segaran 2007](#)). Die Analyse von sozialen Netzwerken hat auch neue Erkenntnisse und Algorithmen im Bereich der sogenannten Collective Intelligence hervorgebracht ([Segaran, Evans & Taylor 2009](#)), auf die sich viele Vertreter als Grundlage für Citizen Science-Projekte berufen. Im Bereich der Digital Humanities hat sich vor allem die Computerlinguistik mit semantischen Konzepten befasst ([Helbig 2001](#)), während sich die Informatik immer mehr mit der Frage beschäftigt, wie qualitativ komplexe Fragestellungen von Computersystemen verarbeitet werden können. Deren Stärke liegt eher auf der quantitativen Verarbeitung von Daten, ungenaue und widersprüchliche Informationen, wie sie in der Geschichtswissenschaft an der Tagesordnung sind, stellen die Informatik vor große Herausforderungen.

Das Konzept digitaler historischer semantischer Netzwerke verbindet viele Aspekte der oben genannten Theorien. Auf formaler Ebene besteht ein solches Netzwerk aus Knoten, welche über Kanten, also Verknüpfungen oder Linien, verbunden sind – ein sogenannter annotierter Multigraph (Diestel 2012). Die Kanten besitzen eine Richtung, was graphisch meist mit Hilfe von Pfeilen dargestellt wird. Das Netzwerk ist deshalb ein Multigraph, weil beliebig viele Kanten dieselben Knoten verbinden können. Außerdem sind sowohl Knoten als auch Kanten mit zusätzlichen Informationen, etwa zu Referenzen, zugeordnete Schlagworte bzw. Tags, Quellenangaben und Verlinkungen auf externe Daten oder Weblinks ausgezeichnet. Anzumerken ist, dass die Kanten zwar gerichtet sind, aber auch in die Gegenrichtung gelesen werden können. Formal sind diese Auszeichnungen den Kanten oder Knoten zugeordnet (Kalus 2010, S. 36 f.). Knoten stellen beliebige Informationseinheiten dar, also beispielsweise eine Person oder Gruppe, ein Ort, eine Region, ein Schriftstück oder auch eine Idee. Kanten verbinden diese Knoten semantisch. Zwei Personen können durch Freundschaft und/oder Verwandtschaft verbunden sein. Eine Person mag als Träger einer Idee gelten, die an einem bestimmten Ort erfunden wurde. In solchen semantischen Netzwerken können sowohl Kanten als auch Knoten mit Zeitpunkten und Zeiträumen ausgezeichnet werden. Da historische semantische Netzwerke über Graphen abgebildet werden, lassen sich Netzwerk- und Graphmethoden auf diese anwenden. Zu diesen gehören Informationen zu Zentralität, Dichte oder das Erkennen von Clustern oder abgeschotteten Netzwerken innerhalb der Knoten einer entsprechenden Datenbank sowie die Messung von sozialer Nähe. Auch Beziehungsstärken können über gewichtete Graphen dargestellt werden. Solche Anwendungen für historische semantische Netzwerke verstehen sich allerdings nicht nur als Auswertungstools. Vielmehr geht es um die Erfassung historischer Daten. Eine der großen Herausforderungen der historischen Forschung ist das Fehlen digitaler, quantitativ auszuwertender Daten. Selbst serielle Quellen müssen oftmals noch von Hand erfasst und ausgewertet werden. Forscher müssen qualitativ komplexe Informationen dafür zunächst händisch in kleinere Wissenseinheiten aufteilen. Ein solches Vorgehen ist nicht neu, wie das Beispiel von Regesten, also die Zusammenfassung von Urkunden oder anderen Quellen zeigt. Die oben genannte Methode geht jedoch weiter und erlaubt das Verknüpfen beliebiger Informationen bereits während des Bearbeitungsprozesses von Quellen. Dies erleichtert die spätere

Auswertung enorm, da auf extrahiertes Wissen schnell zugegriffen werden und dieses bereits weiterführende Informationen enthalten kann. Der einzelne Forscher oder eine Gruppe können dadurch deutlich leichter ihr Wissen sammeln und langfristig für spätere Zwecke sichern.

Ein abschließendes Beispiel soll das gerade vorgestellte Konzept veranschaulichen. In Abbildung 1 ist ein historisches semantisches Netzwerk zu sehen. In diesem existieren diverse Knoten, welche durch Namen und teilweise durch zeitliche Angaben ausgezeichnet sind (Beschreibungen und andere Angaben werden aus Gründen der Übersichtlichkeit nicht dargestellt). Die Symbole zeigen die Art des Knotens an, wie in der Legende beschrieben. Die Abbildung verdeutlicht, dass historische semantische Netzwerke von Menschen leicht interpretiert werden können. Es ist für den Betrachter leicht ersichtlich, dass Ferdinand Kron als ein solcher Knoten eine Person war, die von 1559 bis 1637 gelebt hat. Die Knoten sind semantisch untereinander verknüpft, wie die Beschriftungen über diesen belegen. Auch die Kanten sind teilweise mit zeitlichen Angaben annotiert. Anzumerken ist, dass die Kanten zwar gerichtet sind, aber auch in die Gegenrichtung gelesen werden können. Dabei kann „Ferdinand Kron arbeitet für die Georg Fuggerischen Erben“ auch als „Die Georg Fuggerischen Erben sind die Arbeitgeber von Ferdinand Kron“ interpretiert werden.

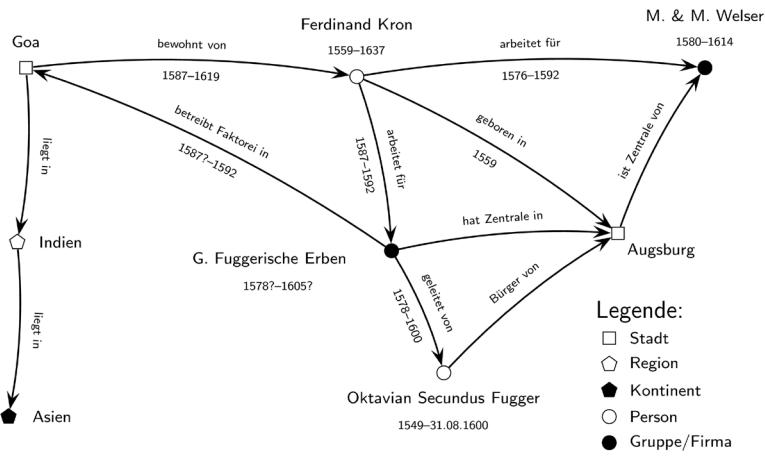


Abbildung 1: Beispiel eines historischen semantischen Netzwerks

Die Abbildung verdeutlicht, dass historische semantische Netzwerke von Menschen leicht interpretiert werden können. Ferdinand Kron stammte also aus Augsburg und lebte lange Zeit in Goa (Indien). Er arbeitete für zwei Augsburger Firmen, zum Teil zur selben Zeit. Zugleich sind die Daten auch für Maschinen leicht zu verarbeiten. Teilgraphen lassen sich so beispielsweise zur Analyse personeller oder geographischer Netzwerke verwenden.

Segrada, eine semantische Graphdatenbank

Auf Basis der oben erläuterten Annahmen hat der Autor dieses Artikels Segrada entwickelt, eine semantische Graphdatenbank. Mit ihrer Hilfe ist es möglich, beliebige Informationen semantisch zu erfassen und untereinander zu verknüpfen. Kern der Anwendung ist eine Webapplikation, die in jedem modernen Browser aufgerufen werden kann. Im Hintergrund läuft eine Java-Anwendung, die auf jedem Betriebssystem mit Java 8 oder höher ausgeführt werden kann. Zum Ausführen sind keine weiteren Vorbereitungen oder besondere Benutzerrechte notwendig. Segrada läuft in einem vom Benutzer änderbaren Ordner mit einem einfach zu bedienenden Startprogramm. Alternativ kann sie auch automatisch im Hintergrund gestartet oder in einem entsprechenden Serverumfeld (als Servlet oder Docker-Instanz) installiert werden (s. Dokumentation auf <http://auxc.de/19c>). Segrada kann also von Einzelpersonen auf dem Desktop verwendet werden, aber auch in Gruppen, die auf einen gemeinsamen Server zugreifen. Anwendung und Datenbank sind zudem auf mehrere Server aufteilbar. Dadurch ist der Größe der Datenbank (fast) kein Limit gesetzt.

Segrada versteht sich als Werkzeug für die wissenschaftliche Arbeit: In der momentanen Version können Knoten und Kanten erstellt und verknüpft werden. Beide Elemente enthalten eine Beschreibung und können mit Schlagwörtern (Tags), Farben und Piktogrammen (Icons) ausgezeichnet werden. Knoten besitzen Titel und Titel-Alternativen, während Verknüpfungen einen Verknüpfungstypen erhalten, der die Art der Verbindung semantisch beschreibt (also beispielsweise „ist Bruder von“). Weitere Auszeichnungen sind möglich: Elemente lassen sich mit beliebig vielen Zeitangaben erweitern, die entweder für einen Zeitpunkt oder

einen Zeitraum stehen. Zeiten können als ungefähre Angaben gekennzeichnet werden und müssen nicht vollständig sein (es genügt z.B. die Angabe eines Jahres). Außerdem können in der aktuellen Version Ortsangaben für Elemente definiert werden. Knoten und Verknüpfungen können mit Quellenverweisen versehen werden, um die Daten wissenschaftlich zu verorten. Zusätzlich lassen sich beliebige Dateien in die Datenbank laden, in den Formaten Microsoft-Word, Open-Office, PDF und Ähnlichem kann der Volltext dieser Dateien erfasst werden, um diesen später nach Begriffen zu durchsuchen. Die Schlagwörter können in einer Hierarchie angeordnet werden, sodass sich komplexe Ordnungssysteme abbilden lassen.

Die Recherche- und Auswertungsmöglichkeiten umfassen im Moment eine Volltext- und mehrere Filtersuchen. Die Daten können außerdem als Tabellen und in einer graphischen Ansicht angezeigt werden, die es erlaubt, Netzwerke genauer zu untersuchen. Noch nicht implementiert sind in der aktuellen Version Möglichkeiten zur Netzwerkanalyse. Diese Funktionen sollen zu einem späteren Zeitpunkt ergänzt werden, ebenso wie automatisierte Vorschläge bei Verknüpfungen. Segrada wird als Open Source unter der Apache-Lizenz veröffentlicht. Fertige Downloads stehen unter <http://segrada.de/> bereit, der Quellcode ist auf Github veröffentlicht. Die Anwendung ist mehrsprachig und unterstützt im Moment Englisch, Deutsch und Rumänisch. Abschließend anzumerken ist, dass Segrada noch als Beta-Version deklariert ist. Zwar sind alle oben genannten Funktionen bereits enthalten, weitere Elemente sollen aber noch hinzukommen. Für den Bereich der Bürgerwissenschaft dürfte es interessant sein, Daten gemeinschaftlich sammeln und innerhalb der Gruppe der beteiligten (Bürger-)Forscher prüfen und teilen zu können. Interessierte Wissenschaftler und Bürger können gleichermaßen die Rohdaten einsehen und entsprechend ihren Forschungsfragen auswerten, sich darüber hinaus aber auch an der Erweiterung der Wissensbasis beteiligen.

Kurze Einführung in die Nutzung von Segrada

Im Folgenden wird die Nutzung von Segrada kurz beschrieben. Zunächst ist es notwendig, eine Java-Laufzeitumgebung auf dem System zu installieren. Außerdem muss Segrada selbst heruntergeladen werden. Das Pa-

ket kann man in einen Ordner seiner Wahl extrahieren, der durch den normalen Desktop-Benutzer geändert werden kann (also z. B. „Dokumente“ oder ein anderer Ordner in „Eigene Dateien“ bzw. im Home-Verzeichnis). Im Anschluss lässt sich Segrada direkt starten. Das Startprogramm dient zur einfachen Kontrolle der Applikation: Standardmäßig wird die Datenbank unter dem Namen `segrada_data` im selben Verzeichnis wie Segrada abgelegt. Mit dem Knopf „Datenbankverzeichnis“ kann dies geändert werden. So lassen sich beispielsweise verschiedene Datenbanken nebeneinander betreiben. Ein Klick auf den Knopf „Starten“ startet die Applikation im Hintergrund. Am Ende der Installation aktiviert sich der dritte Knopf: „Öffne Browser“. Der Standardbrowser öffnet die Adresse `http://localhost:8080/`, unter der Segrada standardmäßig läuft.

Auf dem Bildschirm ist nun das Menü zu sehen. Die Datenbank besitzt standardmäßig noch keine Inhalte. Um einen Wissensknoten zu erstellen, klickt man zunächst auf das „Plus“-Symbol in der Navigationsleiste und wählt den Punkt „Knoten“ aus. Nun erscheint ein Formular zur Eingabe des Datensatzes (siehe Abbildung 2). Einziges Pflichtfeld ist der Titel (mit rotem Stern gekennzeichnet). Weitere Felder sind „Weitere Titel“ und „Beschreibung“.

Zur Auszeichnung des Datensatzes kann dieser mit einer Farbe und einem Piktogramm (Icon) markiert werden, die ihn visuell hervorheben. Farben und eigene Piktogramme können jederzeit selbst definiert bzw. in die Datenbank hoch-

The screenshot shows a web-based form titled "Neuen Knoten erstellen" (Create New Node). The form has the following elements:

- Titel ***: A required text input field with the placeholder "Titel".
- Weitere Titel**: An optional text input field with the placeholder "Weitere Titel".
- Beschreibung**: A larger text input field with the placeholder "Beschreibung".
- Farbe**: A color selection interface showing a grid of 16 color swatches.
- Piktogramm**: A button labeled "Piktogramm wählen" (Choose Pictogram).
- Tags**: A text input field with the instruction "Namen des Tags eintragen und Enter drücken, um diesen hinzuzufügen." (Enter tag names and press Enter to add them).
- Erstellen**: A blue button to submit the form and create the node.

Abbildung 2: Formular zum Anlegen eines neuen Knotens

geladen werden. Zudem können beliebig viele Tags (Stichwörter) definiert werden, bereits definierte Tags werden vom System vorgeschlagen. Um den gewünschten Knoten anzulegen, klickt man auf „Erstellen“. Nach dem Erstellen oder Ändern eines Wissensknotens wird seine Detailansicht eingeblendet. Hier lassen sich Zeit- und Ortsangaben zum Datensatz hinzufügen. Verknüpfungen, Quellen und Dateien zeigen mit diesem Knoten verknüpfte Datensätze. Hier können auch neue Referenzen erstellt werden. Das Eingeben neuer Daten ist folglich relativ leicht mit Segrada. Die Hürde zum Erstellen neuer Daten wurde bewusst niedrig gehalten, um Nutzern schnelle Erfolgserlebnisse zu gewähren. Will man die eingegebenen Knoten anzeigen lassen, wählt man die Listenansicht der Knoten aus (siehe Abbildung 3). Die Listenansicht bietet die Möglichkeit, Knoten nach bestimmten Kriterien zu sortieren und zu filtern. Alternativ kann auch die Volltextsuche verwendet werden, um bestimmte Elemente in der Datenbank wieder zu finden.

Will man die oben erstellten Knoten miteinander semantisch verknüpfen, muss man zunächst Verknüpfungstypen erstellen. Diese definieren die Art der Verknüpfung, also beispielsweise den Verwandtschaftsgrad, eine geographische Verbindung (z. B. Geburtsort) oder Ähnliches. Um einen neuen Verknüpfungstypen anzulegen, wählt man in der Navigationsleiste „Verknüpfungstyp“ aus und definiert diesen über das erscheinende Formular (siehe Abbildung 4). Verknüpfungen sind bidirektional, daher muss man einen Titel für jede Richtung angeben. Beide Titel können gleich sein, beispielsweise bei Freundschaften. Ähnlich wie Knoten besitzen Verknüpfungstypen ein Beschreibungsfeld sowie Farben und Tags.

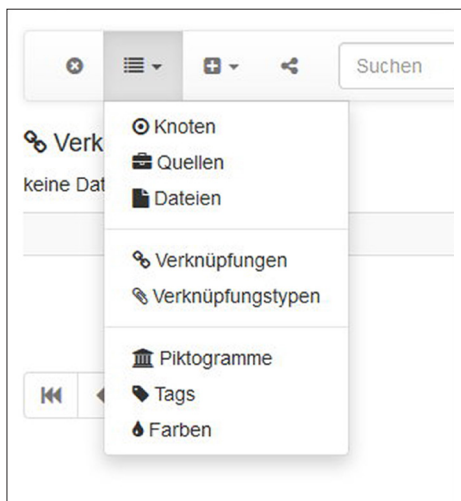


Abbildung 3: Liste der Knoten wählen

Neuen Verknüpfungstyp erstellen

Titel

A → B *

Titel

B → A *

Erlaubte Typen

A → B

Person x

Namen des Tags eintragen und Enter drücken, um diesen hinzuzufügen.

Erlaubte Typen

B → A

Person x

Namen des Tags eintragen und Enter drücken, um diesen hinzuzufügen.

Beschreibung

Beschreibung

Farbe

Piktogramm

Piktogramm wählen

Tags

Abbildung 4: Verknüpfungstyp erstellen

Das Verbinden von Knoten ist ebenfalls relativ einfach. Ähnlich wie oben wählt man aus dem Menü den Punkt „Verknüpfung“ aus. Am besten wählt man zunächst den Verknüpfungstyp aus der Liste aus. Um den „Von“-Knoten zu bestimmen, gibt man seinen Titel oder einen der alternativen Titel ein und wählt in der Liste den gewünschten Treffer aus. Auch hier lässt sich der Datensatz weiter beschreiben und annotieren. Klickt man auf „Erstellen“, gelangt man wieder auf die Detailansicht des Datensatzes (siehe Abbildung 5). Verknüpfungen können mit Datumsangaben und geographischen Koordinaten genauer gekennzeichnet werden.

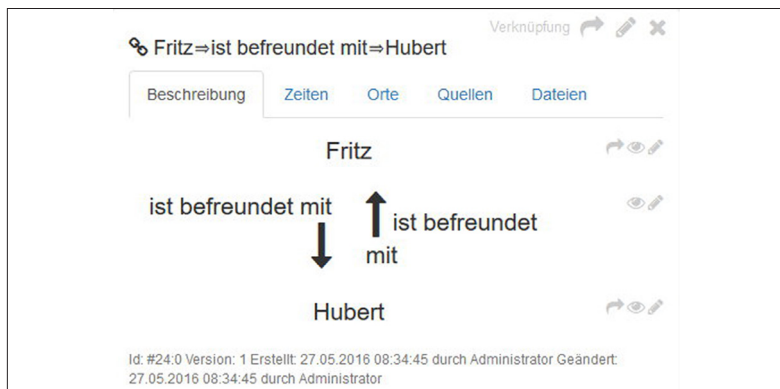


Abbildung 5: Detailansicht einer Verknüpfung

Will man Verknüpfungen graphisch auswerten, kann man in der Detailansicht auf das graue Pfeilsymbol in der oberen rechten Ecke klicken. Dies öffnet die graphische Ansicht, wie in Abbildung 6 zu sehen ist. In der Netzwerk-Ansicht lassen sich Elemente anklicken und verschieben, ein Doppelklick auf einen Knoten öffnet diesen. Damit sind die ersten Schritte getan. Segrada bietet eine ganze Reihe weiterer Möglichkeiten, wie das Erstellen von Quellen, das Hochladen und Indexieren von Dokumenten und verschiedene Suchmöglichkeiten. Auch dies ist in der Onlinehilfe beschrieben.

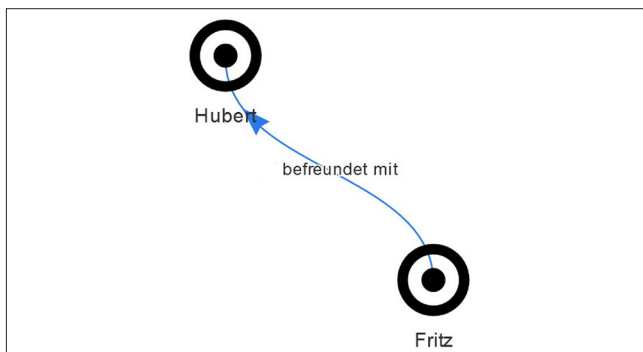


Abbildung 6: Ansicht des semantischen Graphen

Fazit

Das vorgestellte Konzept der historischen semantischen Datenbanken und der Datenbank Segrada zeigen dessen Möglichkeiten. Segrada versteht sich als einfaches Tool für die Forschung und Informationsverarbeitung. Es dürfte nicht nur für Historiker und Bürgerforscher interessant sein, sondern kann auch für andere Einsatzzwecke verwendet werden, wie beispielsweise die Genealogie oder die Erfassung von Figuren in einem Roman. Durch die kostenlose Veröffentlichung steht die Anwendung auch jedem Bürger-Forscher zur Verfügung und ist sofort einsatzbereit. Segrada steht dabei gleichwohl noch am Anfang der Umsetzung, eine zukünftige Weiterentwicklung im Rahmen anderer Projekte ist geplant. Ferner wird Segrada inzwischen von einer kleinen, aber wachsenden Community verwendet, was die Hoffnung nährt, dass das Tool und die Methode weitere spannende Projekte begleiten werden.

Referenzen

Collar, A. (2013): *Religious networks in the Roman Empire*. The spread of new ideas. New York: Cambridge University Press.

Diestel, R. (2012): *Graphentheorie*. (4. Aufl., Nachdruck von 2010), Heidelberg: Springer.

Gramsch, R. (2013): *Das Reich als Netzwerk der Fürsten*. Politische Strukturen unter dem Doppelkönigtum Friedrichs II. und Heinrichs (VII.) 1225–1235. Ostfildern: Thorbecke.

Häberlein, M. (1998): *Brüder, Freunde und Betrüger*. Soziale Beziehungen, Normen und Konflikte in der Augsburger Kaufmannschaft um die Mitte des 16. Jahrhunderts. Colloquia Augustana, 9. Berlin: Akademie Verlag.

Helbig, H. (2001): *Die semantische Struktur natürlicher Sprache*. Wissensrepräsentation mit MultiNet. Berlin: Springer.

Kalus, M. (2010): *Kupfer – Pfeffer – Nachrichten*. Kaufmannsnetzwerke und Handelsstrukturen im europäisch-asiatischen Handel am Ende des 16. Jahrhunderts. Materialien zur Geschichte der Fugger, 6. Augsburg: Wißner.

Kirchner, J. (1901/02): *Prosopographia Attica*. 2 Bde., Berlin: Georg Reimer.

Preiser-Kapeller, J. (2016): *Byzantium's connected Empire, 1282–1402*. A Global History. Im Druck.

Reinhard, W. (1979): *Freunde und Kreaturen*. „Verflechtung“ als Konzept zur Erforschung historischer Führungsgruppen. Römische Oligarchie um 1600. Schriften der Philosophischen Fachbereiche der Universität Augsburg, 14. München: Ernst Vögel.

ders. (Hg., 1996): *Augsburger Eliten des 16. Jahrhunderts*. Prosopographie wirtschaftlicher und politischer Führungsgruppen 1500–1620. Berlin: Akademie Verlag.

Segaran, T. (2007): *Programming Collective Intelligence*. Building Smart Web 2.0 Applications. Sebastopol: O'Reilly.

Segaran, T.; Evans, C.; Taylor, J. (2009): *Programming the Semantic Web*. Sebastopol: O'Reilly.